Research Report

Continuous Exponential Families: An Equating Tool

Shelby J. Haberman

February 2008

ETS RR-08-05

Continuous Exponential Families: An Equating Tool

Shelby J. Haberman
ETS, Princeton, NJ

February 2008

As part of its educational and social mission and in fulfilling the organization's nonprofit charter and bylaws, ETS has and continues to learn from and also to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

ETS Research Reports provide preliminary and limited dissemination of ETS research prior to publication. To obtain a PDF or a print copy of a report, please visit:

http://www.ets.org/research/contact.html

Abstract

Continuous exponential families may be employed to find continuous distributions with the same initial moments as the discrete distributions encountered in typical applications of classical equating. These continuous distributions provide distribution functions and quantile functions that may be employed in equating. To illustrate, an application is considered for a randomly equivalent groups design.

Key words: Moments, information theory

Acknowledgments

I thank Sandip Sinharay, Alina von Davier, and Dan Eignor for their helpful comments.

1 Introduction

In common equipercentile equating methods such as the percentile rank method or kernel equating (von Davier, Holland, & Thayer, 2004), discrete distributions of test scores are approximated by continuous distributions with positive density functions on intervals that include all possible scores. The approximations used are not entirely satisfactory in terms of the relationships of the moments of the approximating distributions and the moments of the original distributions. In addition, use of percentile rank typically results in conversion functions that are not differentiable at all points, while typical use of the kernel method requires both estimation of probabilities by use of log-linear models and smoothing of the resulting distribution function by use of a kernel. One method to reduce this difficulty involves use of continuous exponential families. With continuous exponential families, a one-step construction of a distribution function is provided by a method comparable computationally to use of a log-linear model, and moments are fit exactly where desired. This report describes use of continuous exponential families in equating, develops appropriate methods for estimation and model evaluation, and compares results to those from more conventional approaches to equipercentile equating.

For simplicity, an equivalent groups design is considered in which Test Forms 1 and 2 are compared. Raw scores on Form 1 are integers from c_1 to d_1 and raw scores on Form 2 are integers from c_2 to d_2. For j equals 1 or 2, let X_j be a random variable that represents the score on Form j of a randomly selected population member, so that X_j has integer values from c_j to $d_j > c_j$. To simplify discussion further, assume that, for any integer x from c_j to d_j, X_j equals x with probability $p_j(x) > 0$.

Let F_j denote the distribution function of X_j, so that $F_j(x)$ is the probability that $X_j \leq x$, and let the quantile function Q_j be defined for p in $(0, 1)$ as the smallest x such that $F_j(x) \geq p$. The functions F_j and Q_j are nondecreasing but not continuous, so that they are not readily employed in equating. Instead, equipercentile equating uses continuous random variables A_j such that each A_j has a positive density g_j on an open interval B_j that includes $[c_j, d_j]$, and the distribution function G_j of A_j approximates the distribution function F_j. Because the distribution function G_j is continuous and strictly increasing, the quantile function R_j of A_j is determined by the equation $G_j(R_j(p)) = p$ for p in $(0, 1)$, so that R_j is the inverse G_j^{-1} of G_j. The advantage of R_j over Q_j is that R_j is strictly monotone and continuous. The equating function e_{12} for conversion of a score on Form 1 to a score on Form 2 is then $e_{12}(x) = R_2(G_1(x))$ for x

in B_1, while the equating function e_{21} for conversion of a score on Form 2 to a score on Form 1 is $e_{21}(x) = R_1(G_2(x))$ for x in B_2. Both e_{12} and e_{21} are strictly increasing and continuous on their respective ranges, and e_{12} and e_{21} are inverses, so that $e_{12}(e_{21}(x)) = x$ for x in B_2 and $e_{21}(e_{12}(x)) = x$ for x in B_1. If g_1 is continuous at x in B_1 and g_2 is continuous at $e_{12}(x)$, then application of standard results from calculus shows that the following results hold:

1. At x, the distribution function G_1 is continuously differentiable and has derivative $g_1(x)$.

2. At $e_{12}(x)$, R_2 is continuously differentiable and has derivative $1/g_2(e_{12}(x))$.

3. At x, $e_{12} = R_2(G_1)$ has derivative $e'_{12}(x) = g_1(x)/g_2(e_{12}(x))$.

Similarly, if g_2 is continuous at x in B_2 and g_1 is continuous at $e_{21}(x)$, then e_{21} has derivative $e'_{21}(x) = g_2(x)/g_1(e_{21}(x))$ at x.

1.1 The Percentile-Rank Method

In the percentile-rank method, the distribution of X_1 and X_2 is approximated with the aid of uniformly distributed random variables U_1 and U_2 such that U_1 and X_1 are independent and U_2 and X_2 are independent. The variables U_1 and U_2 have range $(-1/2, 1/2)$. The approximating variable A_j associated with X_j is $X_j + U_j$. If for real x, $[x]$ is the largest integer not greater than x, then a density g_j of A_j may be defined so that $g_j(x) = p_j([x + 1/2])$ for real x in $B_j = (c_j - 1/2, d_j + 1/2)$. For x in B_j,

$$G_j(x) = ([x + 1/2] + 1/2 - x)F_j([x - 1/2]) + (x + 1/2 - [x + 1/2])F_j([x + 1/2]). \tag{1}$$

The functions e_{21} and e_{12} are not differentiable at all points in typical situations, for, in typical cases, g_j is not continuous at integers in $[c_j, d_j]$. One added limitation of the percentile rank method is that the expected value $E(A_j)$ of A_j is the same as the expected value $E(X_j)$ of X_j, but the variance $\sigma^2(A_j)$ of A_j is $\sigma^2(X_j) + 1/12$, a value always greater than the variance $\sigma^2(X_j)$ of X_j.

1.2 General Kernel Equating

In general kernel equating, A_j is constructed so that $E(A_j) = E(X_j)$ and $\sigma^2(A_j) = \sigma^2(X_j)$. Consider continuous independent random variables W_j with common mean 0 and with respective finite variances $\sigma^2(W_j) > 0$. In typical cases, W_j has a normal distribution, but W_j may have a logistic or uniform distribution (Lee & von Davier, in press). As in the percentile rank method, assume that W_j and X_j are independent, assume that each W_j has positive density w_j on a nonempty open interval C_j that includes $(-1/2, 1/2)$, and assume that the W_j are independent of X_1 and X_2. Then the sum $S_j = X_j + W_j$ has a continuous density

$$g_{Sj}(s) = E(w_j(s - X_j)) = \sum_{x=c_j}^{d_j} p_j(x) w_j(s - x) \tag{2}$$

that is positive on an open interval that includes $(c_j - 1/2, d_j + 1/2)$. The variable S_j is the same as the variable A_j in section 1.1 if $W_j = U_j$. The expected value of S_j is $E(S_j) = E(X_j)$, but the variance $\sigma^2(S_j) = \sigma^2(X_j) + \sigma^2(W_j)$ exceeds $\sigma^2(X_j)$.

A linear transformation of the variable S_j is used in kernel equating to provide a new continuous random variable with the same mean and variance as the original variable X_j. Let $\zeta_j = \sigma(X_j)/\sigma(S_j)$, and let $A_j = E(X_j) + \zeta_j[S_j - E(X_j)]$. Then the expectation $E(A_j) = E(X_j)$, and the variance $\sigma^2(A_j) = \sigma^2(X_j)$. Standard rules for calculation of a density under a linear transformation and (2) imply that the random variable A_j is continuous with a density

$$g_{Wj}(x) = \zeta_j^{-1} g_{Sj}(E(X_j) + \zeta_j^{-1}[x - E(X_j)]) = \zeta_j^{-1} \sum_{t=c_j}^{d_j} p_j(t) w_j(\zeta_j^{-1}[x - E(X_j)] - [t - E(X_j)]) \tag{3}$$

that is positive for all x such that $x = E(X_j) + \zeta_j[t - E(X_j)] + y$ for some integer t from c_j to d_j and some y in C_j. The requirement that g_{Wj} be positive on an open interval that includes $[c_j, d_j]$ is certainly satisfied if the density w_j is positive on the real line R, so that $C_j = R$. If W_j has cumulative distribution function H_j, then the distribution function of A_j is

$$G_{Wj}(x) = \sum_{t=c_j}^{d_j} p_j(t) H_j(\zeta_j^{-1}[x - E(X_j)] - [t - E(X_j)]). \tag{4}$$

If the w_j are continuous and the densities g_j are positive on $[c_j, d_j]$, then the conversion functions e_{12} and e_{21} are differentiable.

The limitation still remains that A_j and X_j need not have any common moments of order greater than 2. To be sure, if W_0 is a random variable with mean 0 and all finite moments and if $W_j = h_j W_0$ for positive real h_j, then, as h_j approaches 0, the moments of A_j converge to the moments of X_j; however, this convergence is achieved at a significant cost. As h_j approaches 0, $g_{Wj}(x)$ approaches 0 for x not an integer in $[c_j, d_j]$, and $g_{Wj}(x)$ approaches ∞ for x an integer in $[c_j, d_j]$. In typical cases, the derivatives of e_{12} and e_{21} will become very large at some points.

1.3 Continuous Exponential Families

A much more complete solution to the problem of matching moments may be achieved by use of continuous exponential families. Let q_j and r_j be real numbers such that $q_j < c_j$ and $r_j > b_j$. Let $u_{kj}(x)$, $k \geq 0$, be a polynomial of degree k, so that real constants $h_{kk'}$, $0 \leq k' \leq k$, $k \geq 0$, exist such that

$$x^k = \sum_{k'=0}^{k} h_{kk'j} u_{k'j}(x).$$

Let $\mu_{kj}(X_j)$ be the expectation $E(u_{kj}(X_j))$ of $u_{kj}(X_j)$ for $k \geq 1$. Then the kth moment $E(X_j^k)$ of X_j satisfies

$$E(X_j^k) = \sum_{k'=0}^{k} h_{kk'j} \mu_{k'j}(X_j).$$

Let K be some positive integer, and let $\mathbf{u}_{Kj}(x)$ be the K-dimensional vector with coordinates $u_{kj}(x)$, $1 \leq k \leq K$. Let $\boldsymbol{\mu}_{Kk}(X_j)$ be the K-dimensional vector with coordinates $\mu_{kj}(X_j)$ for $1 \leq k \leq K$. For K-dimensional vectors $\mathbf{x}$ and $\mathbf{y}$ with respective coordinates x_k and y_k, $1 \leq k \leq K$, let $\mathbf{x}'\mathbf{y}$ be the summation $\sum_{k=1}^{K} x_k y_k$. For any K-dimensional vector $\boldsymbol{\theta}$ with coordinates θ_k, $1 \leq k \leq K$, a density $g_{Kj}(\cdot, \boldsymbol{\theta})$ may be defined for x in $[q_j, r_j]$ so that

$$g_{Kj}(x, \boldsymbol{\theta}) = \gamma_{Kj}(\boldsymbol{\theta}) \exp[\boldsymbol{\theta}' \mathbf{u}_{Kj}(x)], \tag{5}$$

where

$$[\gamma_{Kj}(\boldsymbol{\theta})]^{-1} = \int_{q_j}^{r_j} \exp[\boldsymbol{\theta}' \mathbf{u}_{Kj}(x)] dx. \tag{6}$$

In the case of $K = 1$, (5) implies that $g_{Kj}(x, \boldsymbol{\theta})$ is the conditional density of an exponential random variable given that the variable having value between q_j and r_j. For $K = 2$, $g_{Kj}(\boldsymbol{\theta})$ is the conditional density of a normal random variable given that the variable has value between q_j and r_j.

As in Gilula and Haberman (2000), the quality of the approximation provided by the density $g_{Kj}(\cdot, \boldsymbol{\theta})$ in (5) may be assessed by use of the expected logarithmic penalty

$$H_{Kj}(\boldsymbol{\theta}) = -E(\log g_{Kj}(X, \boldsymbol{\theta})) = -\log \gamma_{Kj}(\boldsymbol{\theta}) - \boldsymbol{\theta}' \boldsymbol{\mu}_{Kj}(X_j). \tag{7}$$

The smaller the value of $H_{Kj}(\boldsymbol{\theta})$, the better is the approximation.

Several rationales can be considered for use of the expected logarithmic penalty $H_{Kj}(\boldsymbol{\theta})$ (Gilula & Haberman, 2000). If Y is a continuous real variable with density f, if g is also a probability density function, and if a penalty of $-\log g(y)$ is recorded if $Y = y$, then the smallest expected log penalty $E(-\log g(Y))$ is obtained only if $g(Y) = f(Y)$ with probability 1. This feature in which the penalty is determined by the value of the density at the observed value of Y and the expected penalty is minimized by selection of the actual density is only encountered if the penalty is of the form $a - b\log g(y)$ for $Y = y$ for some real constants a and $b > 0$ such that $b > 0$.

This rationale is not applicable to the discrete variables X_j. In general, if Y is discrete, then the smallest possible expected log penalty $E(-\log g(Y))$ is $-\infty$, for, given any real $c > 0$, g can be chosen so that $g(Y) = c$ with probability 1 and the expected log penalty is $-\log c$. The constant c may be arbitrarily large, so that the expected log penalty may be arbitrarily small. Nonetheless, the criterion $E(-\log g(Y))$ for a density g cannot be made arbitrarily small if adequate constraints are imposed on g. In this section, the requirement that the density function used for prediction of X_j belongs to a continuous exponential family suffices to ensure that, in (7), there is a finite infimum I_{Kj} of the expected log penalty $H_{Kj}(\boldsymbol{\theta})$ over all $\boldsymbol{\theta}$.

A unique K-dimensional vector $\boldsymbol{\theta}_{Kj}$ with coordinates θ_{Kkj}, $1 \le k \le K$, exists such that $H_{Kj}(\boldsymbol{\theta}_{Kj}) = I_{Kj}$. This vector $\boldsymbol{\theta}_{Kj}$ is the unique solution of the equations

$$\nu_{kj}(\boldsymbol{\theta}_K) = \int_{c_j}^{d_j} u_{kj}(x) g_{Kj}(x, \boldsymbol{\theta}_K)dx = \mu_{kj}(X_j),\ 1 \le k \le K. \tag{8}$$

If $\boldsymbol{\nu}_{Kj}(\boldsymbol{\theta}_K)$ is the K-dimensional vector with coordinates $\nu_{kj}(\boldsymbol{\theta}_K)$ for $1 \le k \le K$ and $\boldsymbol{\mu}_{Kj}(X_j)$ is the K-dimensional vector with coordinates $\mu_{kj}(X_j)$ for $1 \le k \le K$, then $\boldsymbol{\nu}_{Kj}(\boldsymbol{\theta}_K) = \boldsymbol{\mu}_{Kj}(X_j)$. Equivalently, if V_{Kj} is a random variable with range $[q_j, r_j]$ with density $g_{Kj}(\cdot, \boldsymbol{\theta}_K)$, then $\mu_{kj}(V_{Kj}) = \mu_{kj}(X_j)$ for all integers from 1 to K, so that $E(V_{Kj}^k) = E(X_j^k)$ for $1 \le k \le K$. If $K \ge 1$, then $E(V_{Kj}) = E(X_j)$. If $K \ge 2$, then $\sigma^2(V_{Kj}) = \sigma^2(X_j)$. If $K \ge 3$, then V_{Kj} and X_j have the same skewness coefficient. If $K \ge 4$, then V_{Kj} and X_j have the same coefficient of kurtosis. By (7) and (8), the minimum expected penalty

$$I_{Kj} = -\log \gamma_{Kj}(\boldsymbol{\theta}_{Kj}) - \boldsymbol{\theta}'_{Kj}\boldsymbol{\mu}_{Kj}(X_j). \tag{9}$$

Corresponding to the density $g_{Kj}(\cdot, \boldsymbol{\theta}_{Kj})$ in (5) is the cumulative distribution function

$$G_{Kj}(x) = \int_{q_j}^{x} g_{Kj}(v, \boldsymbol{\theta}_{Kj})dv \tag{10}$$

for x between q_j and r_j. One then has an inverse R_{Kj} such that $G_{Kj}(R_{Kj}(p)) = p$ for p in $(0, 1)$. In equating, a positive integer K_j is selected for each j. Then $e_{12}(x) = R_{K_2 2}(G_{K_1 1}(x))$ for x in (q_1, r_1) and $e_{21}(x) = R_{K_1 1}(G_{K_2 2}(x))$ for x in (q_2, r_2).

In section 2, estimation of $\boldsymbol{\theta}_{Kj}$, I_{Kj}, G_{Kj}, and R_{Kj} is considered for the case of simple random sampling. Estimates, large sample approximations for distributions of estimates, and estimated asymptotic standard deviations are all provided.

In section 3, some examples of estimation are provided for some distributions of test scores reported in von Davier et al. (2004). In section 4, conclusions are reached concerning the status of continuous exponential families in equating.

In sections 2, 3, and 4, comparisons with alternative equating methods are considered. For this purpose, some consideration of expected log penalty for percentile-rank and kernel methods is provided in section 1.4

1.4 Comparisons by Expected Log Penalty

The proposed approximations may be compared to those from percentile-rank or kernel equating. In the percentile-rank case, the expected log penalty

$$I_{Pj} = -E(\log p_j(X_j)) = -\sum_{x=c_j}^{d_j} p_j(x) \log p_j(x) \tag{11}$$

is the entropy of the discrete variable X_j. In the percentile-rank case with log-linear smoothing of order $K \leq d_j - c_j$, the probabilities $p_j(x)$ are approximated by probabilities $p_{Kj}(x)$ defined so that $\log p_{Kj}(x)$ is a polynomial in x of order K and the expected penalty

$$I_{PKj} = -E(\log p_{Kj}(X_j)) = -\sum_{x=c_j}^{d_j} p_j(x) \log p_{Kj}(x) \geq I_{Pj} \tag{12}$$

is minimized subject to this constraint on $p_{Kj}(x)$. One has

$$p_{Kj}(x) = \eta_{Kj}(\boldsymbol{\omega}_{Kj}) \exp[\boldsymbol{\omega}'_{Kj}\mathbf{u}_{Kj}(x)], \tag{13}$$

$$[\eta_{Kj}(\boldsymbol{\omega}_{Kj})]^{-1} = \sum_{x=c_j}^{d_j} \exp[\boldsymbol{\omega}'_{Kj}\mathbf{u}_{Kj}(x)], \tag{14}$$

and

$$\sum_{x=c_j}^{d_j} u_k(x) p_{Kj}(x) = \mu_{kj}(X_j),\ 1 \leq k \leq K. \tag{15}$$

Because a polynomial of degree $d_j - c_j$ can be found to fit any real function at $d_j - c_j$ points, if $K = d_j - c_j$, then $I_{PKj} = I_{Kj}$ and $p_{Kj}(x) = p_j(x)$. In general, the equation $I_{PKj} = I_{Pj}$ holds if, and only if, $\log p_j(x)$ is a polynomial of order K in terms of x, so that the smoothed probability $p_{Kj}(x)$ is equal to the actual probability.

In kernel equating, (3) implies that the expected log penalty is

$$I_{Wj} = -E(\log g_{Wj}(X)) = -\sum_{x=c_j}^{d_j} p_j(x) \log g_{Wj}(x). \tag{16}$$

With log-linear smoothing of order $K \geq 2$, the expected log penalty is

$$I_{WKj} = -\sum_{x=c_j}^{d_j} p_j(x) \log g_{WKj}(x), \tag{17}$$

where

$$g_{WKj}(t) = \zeta_j^{-1} \sum_{x=c_j}^{d_j} p_{Kj}(x) w_j(\zeta_j^{-1}[t - E(X)] - [x - E(X)]). \tag{18}$$

The expected log penalty can be made arbitrarily small by selection of $W_j = h_j W_0$, where h_j is positive, W_0 is a continuous random variable with positive density w_0, and W_0 is independent of X_1 and X_2. Because $w_j(t) = w_0(t/h_j)/h_j$, it follows that, as h_j approaches 0, ζ_j approaches 1 and $h_j g_{WKj}(x)$ has a lower limit at least equal to $p_{Kj}(x) w_0(0)$ for each integer x from c_j to d_j. Thus I_{WKj} approaches $-\infty$. As evident from the formulas in the introduction for conversions, a very large value of g_{WKj} has the danger that the derivative of a conversion will be very large and the conversion will be unstable. This issue is customarily treated in kernel equating (von Davier et al., 2004, pp. 62–64). One relatively simple approach based on the expected log penalty is to require that $g_{WKj}(t)$ have no more than $K-1$ points at which its derivative changes sign, just as $g_{Kj}(t)$ has a derivative that changes sign no more than $K-1$ times.

2 Estimation of Parameters Under Random Sampling

Data from random sampling are readily applied to estimation of the parameters $\boldsymbol{\theta}_K$ for $K \geq 1$ (Gilula & Haberman, 2000). Recall definitions in section 1.3. For j equals 1 or 2, let X_{ij}, $1 \leq i \leq n_j$, be independent and identically distributed random variables with the same distribution as X_j. Let $m_{kj}(X_j)$ be the sample mean

$$m_{kj}(X_j) = n_j^{-1} \sum_{i=1}^{n_j} u_{kj}(X_{ij}) \tag{19}$$

for $k \geq 1$, and let $\mathbf{m}_{Kj}(X_j)$ be the K-dimensional vector with coordinates $m_{kj}(X_j)$ for $1 \leq k \leq K$. If the X_{ij}, $1 \leq i \leq n_j$, have at least K distinct values, then $\boldsymbol{\theta}_{Kj}$ is estimated by the K-dimensional vector $\hat{\boldsymbol{\theta}}_{Kj}$ with coordinates $\hat{\theta}_{Kkj}$, $1 \leq k \leq K$, where $\hat{\boldsymbol{\theta}}_{Kj}$ is the unique K-dimensional vector such that

$$\nu_{kj}(\hat{\boldsymbol{\theta}}_{Kj}) = m_{kj}(X_j),\ 1 \leq k \leq K. \tag{20}$$

Thus (20) corresponds to (8. As the sample size n_j approaches ∞, $\hat{\boldsymbol{\theta}}_{Kj}$ converges to $\boldsymbol{\theta}_{Kj}$ with probability 1, and $n_j^{1/2}(\hat{\boldsymbol{\theta}}_{Kj} - \boldsymbol{\theta}_{Kj})$ converges in distribution to a multivariate normal random variable with zero mean and with covariance matrix $\mathbf{B}_{Kj} = \mathbf{C}_{Kj}^{-1}\mathbf{D}_{Kj}\mathbf{C}_{Kj}^{-1}$. Here $\mathbf{D}_{Kj}$ is the covariance matrix of $\mathbf{u}_{Kj}(X_j)$ and $\mathbf{C}_{Kj}$ is the covariance matrix of the K-dimensional vector $\mathbf{u}_{Kj}(V_{Kj})$. Thus $\mathbf{C}_{Kj} = \mathbf{U}_{Kj}(\boldsymbol{\theta}_{Kj})$, where row k and column k' of $\mathbf{U}_{Kj}(\boldsymbol{\theta}_{Kj})$ is

$$U_{Kkk'j}(\boldsymbol{\theta}_{Kj}) = \int_{c_j}^{d_j} u_{kj}(x)u_{k'j}(x)g_{Kj}(x,\boldsymbol{\theta}_{Kj})dx - \nu_{kj}(\boldsymbol{\theta}_{Kj})\nu_{k'j}(\boldsymbol{\theta}_{Kj}). \tag{21}$$

One may estimate $\mathbf{W}_{Kj}$ by $\hat{\mathbf{C}}_{Kj} = \mathbf{U}_{Kj}(\hat{\boldsymbol{\theta}}_{Kj})$, and $\mathbf{D}_{Kj}$ may be estimated by the sample covariance matrix $\hat{\mathbf{D}}_{Kj}$ of $\mathbf{u}_K(X_j)$. Thus $\mathbf{B}_{Kj}$ is estimated by $\hat{\mathbf{B}}_{Kj} = \hat{\mathbf{C}}_{Kj}^{-1}\hat{\mathbf{D}}_{Kj}\hat{\mathbf{C}}_{Kj}^{-1}$. The estimated asymptotic standard deviation (EASD) of $\hat{\theta}_{Kkj}$ is $\hat{\sigma}(\hat{\theta}_{Kkj}) = (n_j^{-1}\hat{B}_{Kkkj})^{1/2}$, where $\hat{B}_{Kkkj}$ is row k and column k of $\hat{\mathbf{B}}_{Kj}$.

The minimum expected penalty I_{Kj} in (9) may be estimated by

$$\hat{I}_{Kj} = -\log \gamma_{Kj}(\hat{\boldsymbol{\theta}}_{Kj}) - \hat{\boldsymbol{\theta}}_{Kj}'\mathbf{m}_{Kj}(X_j). \tag{22}$$

The estimate $\hat{I}_{Kj}$ of (22) has the standard stability property that, as the sample size n increases, $\hat{I}_{Kj}$ converges to I_{Kj} with probability 1 and $n_j^{1/2}(\hat{I}_{Kj} - I_{Kj})$ converges in distribution to a normal random variable with mean 0 and variance

$$\sigma^2(-\log g_{Kj}(X,\boldsymbol{\theta}_{Kj})) = [\mu_{Kj}(X_j)]'\mathbf{B}_{Kj}\mu_{Kj}(X_j).$$

Let $\hat{p}_j(x)$ be the fraction of observations i from 1 to n_j with $X_{ij} = x$, and let $0 \log 0$ be 0. The EASD of $\hat{I}_{Kj}$ is then

$$\hat{\sigma}(\hat{I}_{Kj}) = (n_j^{-1}[\mathbf{m}_{Kj}(X_j)]'\hat{\mathbf{B}}_{Kj}\mathbf{m}_{Kj})^{1/2}. \tag{23}$$

Equivalently, (23) can be written in terms of the density g_{Kj} of (5). One has

$$\hat{\sigma}(\hat{I}_{Kj}) = \left\{ n_j^{-1} \sum_{x=c_j}^{d_j} [-\hat{p}_j(x) \log g_{Kj}(x, \hat{\boldsymbol{\theta}}_{Kj}) - \hat{I}_{Kj})^2 \right\}^{1/2}. \tag{24}$$

2.1 Alternative Estimation Methods and Information

In comparisons with the alternative methods of section 1.1, 1.2, and 1.4, the expected log penalty I_{Pj} for the percentile-rank case defined by (11) is estimated by

$$\hat{I}_{Pj} = -\sum_{x=c_j}^{d_j} \hat{p}_j(x) \log \hat{p}_j(x), \tag{25}$$

and the EASD of $\hat{I}_{Pj}$ is

$$\hat{\sigma}(\hat{I}_{Pj}) = \left\{ n_j^{-1} \sum_{x=c_j}^{d_j} \hat{p}_j(x)[-\log \hat{p}_j(x) - \hat{I}_{Ej}]^2 \right\}^{1/2}. \tag{26}$$

In the percentile-rank case with log-linear smoothing of order K, the estimated expected log penalty that corresponds to I_{PKj} in (12) is

$$\hat{I}_{PKj} = -\sum_{x=c_j}^{d_j} \hat{p}_j(x) \log \hat{p}_{Kj}(x), \tag{27}$$

where (13), (14), and (15) lead to

$$\hat{p}_{Kj}(x) = \eta_{Kj}(\hat{\boldsymbol{\omega}}_{Kj}) \exp[\hat{\boldsymbol{\omega}}'_{Kj} \mathbf{u}_{Kj}(x)] \tag{28}$$

and

$$\sum_{x=c_j}^{d_j} u_{kj}(x) \hat{p}_{Kj}(x) = m_{kj}(X_j),\ 1 \leq k \leq K.$$

The EASD of $\hat{I}_{PKj}$ is

$$\hat{\sigma}(\hat{I}_{PKj}) = \left\{ n_j^{-1} \sum_{x=c_j}^{d_j} \hat{p}_j(x)[-\log \hat{p}_{Kj}(x) - \hat{I}_{EKj}]^2 \right\}^{1/2}. \tag{29}$$

In kernel equating with a fixed choice of W_j, I_{Wj} in (16) may be estimated by

$$\hat{I}_{Wj} = -n_j^{-1} \sum_{x=c_j}^{d_j} \log \hat{g}_{Wj}(x), \tag{30}$$

where the g_{Wj} of (3) is estimated by

$$\hat{g}_{Wj}(t) = \hat{\zeta}_j^{-1} \sum_{x=c_j}^{d_j} \hat{p}_j(x) w_j(\hat{\zeta}_j^{-1}(t - \bar{X}_j) - [x - \bar{X}_j]), \tag{31}$$

$\bar{X}_j$ is the sample mean $m_1(X_j)$ of the X_{ij}, $1 \le i \le n_j$, $\hat{\sigma}(X_j)$ is the sample standard deviation of the X_i, and ζ_j is estimated by

$$\hat{\zeta}_j = \hat{\sigma}(X_j)/[\hat{\sigma}^2(X_j) + \sigma^2(W_j)]^{1/2}.$$

The formula for the EASD of $\hat{I}_{Wj}$ is somewhat more complicated than in other cases, so it is omitted.

With log-linear smoothing of order $K \ge 2$, the expected log penalty I_{WKj} of (17) is estimated by

$$\hat{I}_{WKj} = -\sum_{x=c_j}^{d_j} \hat{p}_{Kj}(x) \log \hat{g}_{WKj}(x), \tag{32}$$

where g_{WKj} in (18) is estimated by

$$\hat{g}_{WKj}(t) = \hat{\zeta}_j^{-1} \sum_{x=c_j}^{d_j} \hat{p}_{Kj} w_j(\hat{\zeta}_j^{-1}[t - \bar{X}_j] - [x - \bar{(}X)_j]). \tag{33}$$

The EASD of the estimate $\hat{I}_{WKj}$ in (32) is a bit more complex than in the case of the estimate $\hat{I}_{Wj}$ in (30), so this formula is also omitted.

2.2 Equating Functions for Continuous Exponential Families

The distribution function G_{Kj} in (10) for the continuous exponential family has estimate $\hat{G}_{Kj}$ defined by

$$\hat{G}_{Kj}(x) = \int_{q_j}^{x} g_{Kj}(v, \hat{\boldsymbol{\theta}}_{Kj}) dv \tag{34}$$

for $q_j \le x \le r_j$, and the quantile function R_{Kj} corresponding to G_{Kj} has estimate $\hat{R}_{Kj}$ defined by $\hat{G}_{Kj}(\hat{R}_{Kj}(p)) = p$ for $0 < p < 1$. Standard large-sample arguments imply that, as the sample size n_j approaches ∞, $\hat{G}_{Kj}(x)$ converges to $G_{Kj}(x)$ with probability 1 for $q_j \le x \le r_j$, so that $|\hat{G}_{Kj} - G_{Kj}|$, the supremum of $|\hat{G}_K(x) - G_K(x)|$ for $q_j \le x \le r_j$, converges to 0 with probability 1. In addition, $[\hat{G}_{Kj}(x) - F_{Kj}(x)]/\sigma(\hat{G}_{Kj}(x))$ converges in distribution to a normal random variable with mean 0 and variance 1 if the asymptotic standard deviation of $\hat{G}_{Kj}(x)$ is

$$\sigma(\hat{G}_{Kj}(x)) = \{n_j^{-1}\{[\mathbf{T}_{Kj}(x)]'\mathbf{B}_{Kj}\mathbf{T}_{Kj}(x)\}^{1/2} \tag{35}$$

and if

$$\mathbf{T}_{Kj}(x) = \int_{q_j}^{x} [\mathbf{u}_{Kj}(v) - \boldsymbol{\mu}_{Kj}(X_j)] g_{Kj}(v, \boldsymbol{\theta}_{Kj}) dv. \tag{36}$$

Similarly, the estimated quantile function $\hat{R}_{Kj}(p)$ corresponding to the distribution function $\hat{G}_{Kj}$ in (34) converges to the quantile function $R_{Kj}(p)$ with probability 1, and $[\hat{R}_{Kj}(p)-R_{Kj}(p)]/\sigma(\hat{R}_{Kj}(p))$ converges in distribution to a normal random variable with mean 0 and variance 1 if the asymptotic standard deviation of $\hat{R}_{Kj}(p)$ is

$$\sigma(\hat{R}_{Kj}(p)) = n_j^{-1/2}[g_{Kj}(R_{Kj}(p))]^{-1}\sigma(\hat{G}_{Kj}(R_{Kj}(p))). \tag{37}$$

Estimated asymptotic standard deviations may be derived by use of obvious substitutions of estimated parameters for actual parameters. Thus (35) for the asymptotic standard deviation $\hat{R}_{Kj}(p)$ leads to the estimated asymptotic standard deviation

$$\hat{\sigma}(\hat{G}_{Kj}(x)) = \{n_j^{-1}[\hat{\mathbf{T}}_{Kj}(x)]'\hat{\mathbf{B}}_{Kj}\hat{\mathbf{T}}_{Kj}(x)\}^{1/2}, \tag{38}$$

where the vector $\mathbf{T}_{Kj}(x)$ of (36) is estimated by

$$\hat{\mathbf{T}}_{Kj}(x) = \int_{q_j}^{x}[\mathbf{u}_K(v) - \mathbf{m}_K(X_j)]g_{Kj}(v, \hat{\boldsymbol{\theta}}_{Kj})dv, \tag{39}$$

and the (37) for the asymptotic standard deviation of $\hat{R}_{Kj}(p)$ leads to the estimated asymptotic standard deviation

$$\hat{\sigma}(\hat{R}_{Kj}(p)) = n_j^{-1/2}[g_{Kj}(\hat{R}_{Kj}(p))]^{-1}\hat{\sigma}(\hat{G}_{Kj}(R_{Kj}(p))). \tag{40}$$

In the case of equating with constants K_1 for X_1 and K_2 for X_2, the conversion function $e_{12}(x)$ for conversion of Score x on Form 1 to a score on Form 2 has estimate $\hat{e}_{12}(x) = \hat{R}_{K_2 2}(\hat{G}_{K_1 1}(x))$ for $q_1 < x < r_1$, and the conversion function $e_{21}(x)$ for conversion of a Score x on Form 2 to a score on Form 1 has estimate $\hat{e}_{21}(x) = \hat{R}_{K_1 1}(\hat{G}_{K_2 2}(x))$ for $q_2 < x < r_2$. As the sample sizes n_1 and n_2 become large, $\hat{e}_{12}(x)$ converges with probability 1 to $e_{12}(x)$, and $\hat{e}_{21}(x)$ converges with probability 1 to $e_{21}(x)$. In addition, $(\hat{e}_{12} - e_{12})/\sigma(\hat{e}_{12})$ converges in distribution to a standard normal random variable if the asymptotic standard deviation of the estimated conversion $\hat{e}_{12}(x)$ is

$$\sigma(\hat{e}_{12}(x)) = [\sigma^2(\hat{G}_{K_2 2}(e_{12}(x))) + \sigma^2(\hat{G}_{K_1 1}(x))]^{1'2}/g_{K_2 2}(e_{12}(x)). \tag{41}$$

Given (41), it follows that the EASD of the estimated conversion $\hat{e}_{12}(x)$ is

$$\hat{\sigma}(\hat{e}_{12}(x)) = [\hat{\sigma}^2(\hat{G}_{K_2 2}(e_{12}(x))) + \hat{\sigma}^2(\hat{G}_{K_1 1}(x))]^{1/2}/\hat{g}_{K_2 2}(\hat{e}_{12}(x))) \tag{42}$$

The case of the conversion function e_{21} from Form 2 to Form 1 is treated in a similar fashion.

2.3 Other Equating Functions

The large-sample results for equating based on continuous exponential families are a bit simpler than those for percentile ranks or for the kernel method; however, the two cases differ somewhat.

In the case of percentile ranks, for x in the open interval $B_j = (c_j - 1/2, d_j + 1/2)$, the continuous distribution function G_j in (1) may be estimated by

$$\hat{G}_j(x) = ([x + 1/2] + 1/2 - x)\hat{F}_j([x - 1/2]) + (x + 1/2 - [x + 1/2])\hat{F}_j([x + 1/2]), \tag{43}$$

where $\hat{F}_j$ is the empirical distribution function of F_j. With probability 1, $|\hat{G}_j - G_j|$ converges to 0 as n_j becomes large. If, for $0 < p < 1$, the estimated quantile function $\hat{R}_j$ corresponding to $\hat{G}_j$ satisfies $\hat{G}_j(\hat{R}_j(p)) = p$ and the quantile function R_j corresponding to G_j satisfies $G_j(R_j(p)) = p$, then the estimated quantile function $\hat{R}_j(p)$ converges to the quantile function $R_j(p)$ with probability 1, so that the estimated conversion function $\hat{e}_{12}(x) = \hat{R}_2(\hat{G}_1(x))$ for conversion of Score x on Form 1 to a score on Form 2 converges to the corresponding convergence function $e_{12}(x) = R_2(G_1(x))$ for x in B_1. Results for asymptotic normality are not entirely satisfactory, for a case with $e_{12}(x) - 1/2$ equal to an integer typically results in no normal approximation for the distribution of the estimated conversion function $\hat{e}_{12}(x)$. Similar issues arise for the percentile-rank method with log-linear smoothing.

Asymptotic results are available for kernel equating (von Davier et al., 2004) that are comparable to those for continuous exponential families. When kernel equating is applied with log-linear smoothing, the results are a bit more complicated than for continuous exponential families due to the use of both smoothing of frequencies and conversion to a continuous distribution by use of the kernel approach.

2.4 Computational Issues

Given a starting value $\boldsymbol{\theta}_{Kj0}$, the Newton-Raphson algorithm may be employed to compute $\hat{\boldsymbol{\theta}}_{Kj}$ in (20). At step $t \geq 0$, a new approximation $\boldsymbol{\theta}_{Kj(t+1)}$ of $\hat{\boldsymbol{\theta}}_{Kj}$ is found by the equation

$$\boldsymbol{\theta}_{Kj(t+1)} = \boldsymbol{\theta}_{Kjt} + [\mathbf{U}_{Kj}(\boldsymbol{\theta}_{Kjt})]^{-1}[\mathbf{m}_{Kj}(X_j) - \boldsymbol{\nu}_{Kj}(\boldsymbol{\theta}_{Kjt})]. \tag{44}$$

Recall that the elements of the K by K matrix $\mathbf{U}_{Kj}$ are defined in (21), the elements of the K-dimensional vector $\mathbf{m}_{Kj}(X_j)$ are defined in (19), and the elements of the K-dimensional vector function $\boldsymbol{\nu}_{Kj}$ are defined in (8).

In practice, numerical work is simplified if computations employ Legendre polynomials (Abramowitz & Stegun, 1965, chapters 8, 22). The Legendre polynomial of degree 0 is $P_0(x) = 1$, the Legendre polynomial of degree 1 is $P_1(x) = x$, and the Legendre polynomial $P_{k+1}(x)$ of degree $k + 1$, $k \geq 1$, is determined by the recurrence relationship

$$P_{k+1}(x) = (k+1)^{-1}[(2k+1)xP_k(x) - kP_{k-1}(x)], \tag{45}$$

so that $P_2(x) = (3x^2 - 1)/2$. These polynomials satisfy the relationships

$$\int_{-1}^{1} P_j(x)P_k(x)dx = \delta_{jk}\frac{2}{2k+1} \tag{46}$$

for nonnegative integers j and k, where the Kronecker delta δ_{jk} is 1 for $j = k$ and 0 otherwise. It is relatively efficient for numerical work to let $u_{kj}(x) = P_k((2x - q_j - r_j)/(r_j - q_j))$, for then the K by K matrix $\mathbf{U}_{Kj}(\mathbf{0}_K)$ defined in (21) is a diagonal matrix, where $\mathbf{0}_K$ is the K-dimensional vector with all coordinates 0. The obvious choice $u_{kj}(x) = x^k$ is avoided because this case often leads to poor conditioning of the matrix $\mathbf{U}_{Kj}(\boldsymbol{\theta})$. The Legendre polynomials also form the basis for the Gaussian quadratures required for evaluation of the integrals from c_j to d_j that are needed in numerical work (Abramowitz & Stegun, 1965, p. 887). In this paper, calculations use 8-point Gaussian quadrature.

3 Example

Table 7.1 of von Davier et al. (2004) provides two distributions of test scores that are integers from $c_j = 0$ to $d_j = 20$. To illustrate results, the case of $q_j = -0.5$ and $r_j = 20.5$ is considered for K from 2 to 4 and for the Legendre polynomial case with $u_{kj}(x) = P_k((2x - q_j - r_j)/(r_j - q_j))$ for P_k defined as in (45). Results for parameters are summarized in Tables 1 and 2. Results in terms of estimated expected log penalties are summarized in Table 3. These tables suggest that gains over the quadratic case ($K = 2$) are very modest for both X_1 and X_2, although some evidence exists that, for both variables, the parameters θ_{3j3}, θ_{4j3}, and θ_{4j4} of (8) are nonzero. Estimated parameters in Tables 1 and 2 are computed as in section 2.4. The estimated asymptotic standard deviations are found as in section 2. In Table 3, (22) and (23) are employed to obtain estimates.

Use of equipercentile equating leads to somewhat similar results. For X_1, the estimated expected log penalty $\hat{I}_{P1}$ in (25) is 2.741, and the EASD from (26) is 0.015. For X_2, $\hat{I}_{P2}$ is 2.765, and the EASD is 0.014. For the case of $K = 2$, for X_1, the estimated expected log penalty $\hat{I}_{P21}$

Table 1

Parameters for Variable X_1

Parameter	Estimate	EASD
θ_{21}	0.590	0.074
θ_{22}	-2.364	0.097
θ_{31}	0.701	0.100
θ_{32}	-2.415	0.103
θ_{33}	0.172	0.112
θ_{41}	0.792	0.124
θ_{42}	-2.681	0.172
θ_{43}	0.294	0.140
θ_{44}	-0.322	0.150

Note. EASD = estimated asymptotic standard deviation.

Table 2

Parameters for Variable X_2

Parameter	Estimate	EASD
θ_{21}	1.059	0.076
θ_{22}	-2.105	0.094
θ_{31}	1.212	0.110
θ_{32}	-2.224	0.117
θ_{33}	0.231	0.112
θ_{41}	1.287	0.137
θ_{42}	-2.372	0.172
θ_{43}	0.338	0.150
θ_{44}	-0.173	0.132

Note. EASD = estimated asymptotic standard deviation.

Table 3

Estimated Expected Log Penalties for Variables X_1 and X_2

Variable	Degree	Estimate	EASD
X_1	2	2.747	0.015
X_1	3	2.747	0.015
X_1	4	2.745	0.015
X_2	2	2.773	0.014
X_2	3	2.772	0.014
X_2	4	2.771	0.014

Note. EASD = estimated asymptotic standard deviation.

from (27) is 2.748, and the EASD from (29) is 0.015. For X_2, $\hat{I}_{P22}$ is 2.773, and the EASD is 0.014. As one illustration of results for kernel equating, consider the case of W_1, a normal random variable with mean 0 and standard deviation 0.622; W_2, a normal random variable with mean 0 and standard deviation 1.367; and $K = 2$ for both X_1 and X_2 (von Davier et al., 2004, p. 106). Computations for the kernel method are described in the user guide for Version 2.1 of the LOGLIN/KE program (Chen, Yan, Han, & von Davier, 2006). For X_1, the estimated expected penalty $\hat{I}_{W21}$ from (32) is 2.748. For X_2, $\hat{I}_{W22}$ is 2.779. In both cases, the kernel density has a derivative that only changes sign at $K - 1 = 1$ points, so that the criterion of section 1.4 for the kernel density is satisfied. At least for the example under study, it appears that the equipercentile, kernel, and continuous exponential family approaches lead to comparable results in terms of compatibility with the data.

Equating results may now be considered. The case of the conversion e_{12} from Form 1 to Form 2 will be examined for the cases under study. Results are provided in Table 4. They employ formulas developed in sections 2.2 and 2.3. In kernel and percentile-rank equating, log-linear smoothing is used with the constant K equal to 2 for each variable. For continuous exponential families, $K_1 = K_2 = 2$. These results are an illustration of one of a very large number of possibilities. In this example, the three conversions are very similar for all possible values of X_1. For the two methods for which estimated asymptotic standard deviations are available, results are rather similar. The results for the continuous exponential family are relatively best at the extremes of the distribution.

Table 4
Comparison of Conversions From Form 1 to Form 2

	Continuous exponential		Kernel		Percentile rank
Value	Estimate	EASD	Estimate	EASD	Estimate
0	0.091	0.110	-0.061	0.194	0.095
1	1.215	0.209	1.234	0.235	1.179
2	2.304	0.239	2.343	0.253	2.255
3	3.377	0.240	3.413	0.253	3.325
4	4.442	0.230	4.473	0.242	4.392
5	5.504	0.214	5.529	0.225	5.458
6	6.564	0.198	6.582	0.207	6.522
7	7.621	0.182	7.634	0.189	7.585
8	8.677	0.169	8.685	0.174	8.647
9	9.732	0.159	9.734	0.162	9.706
10	10.784	0.155	10.781	0.155	10.761
11	11.834	0.155	11.825	0.153	11.823
12	12.880	0.160	12.865	0.156	12.859
13	13.919	0.168	13.900	0.163	13.898
14	14.950	0.177	14.925	0.172	14.928
15	15.966	0.184	15.936	0.179	15.947
16	16.959	0.187	16.925	0.182	16.949
17	17.912	0.179	17.879	0.178	17.927
18	18.802	0.156	18.799	0.164	18.871
19	19.592	0.109	19.723	0.145	19.760
20	20.240	0.040	20.818	0.119	20.380

Note. EASD = estimated asymptotic standard deviation.

4 Conclusions

Results in this report suggest that equating via continuous exponential families can be regarded as a viable competitor to kernel equating. Continuous exponential families lead to simpler procedures and more thorough moment agreement, for fewer steps are involved in equating by continuous exponential families due to elimination of kernel smoothing. In addition, equating by continuous exponential families does not require selection of bandwidths.

One example does not produce an operational method, and kernel equating is rapidly approaching operational use, so it is important to consider some required steps.

Although equivalent-group designs are used in operations both at ETS and elsewhere, a large fraction of equating designs are more complex and require at least treatment of bivariate distributions. For this purpose, continuous exponential families can be employed, for continuous exponential families can be applied to multivariate distributions. This issue is expected to be the subject of future work. No reason exists to expect that continuous exponential families cannot be applied to any standard equating situation to which kernel equating has been applied.

It is certainly appropriate to consider a variety of applications to data, and some work on quality of large-sample approximations is appropriate when smaller sample sizes are contemplated.

References

Abramowitz, M., & Stegun, I. A. (1965). *Handbook of mathematical functions.* New York: Dover.

Chen, H., Yan, D., Han, N., & von Davier, A. (2006). *LOGLIN/KE user guide: Version 2.1* [Computer software manual]. Princeton, NJ: ETS.

Gilula, Z., & Haberman, S. J. (2000). Density approximation by summary statistics: An information-theoretic approach. *Scandinavian Journal of Statistics*, *27*, 521–534.

Lee, Y.-H., & von Davier, A. (in press). *Comparing alternative kernels for the kernel method of test equating: Gaussian, logistic, and uniform kernels.* Princeton, NJ: ETS.

von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004). *The kernel method of test equating.* New York: Springer.